

### **AMENDMENTS TO THE CLAIMS**

This Listing of Claims will replace all prior versions, and listings, of claims in the Application:

#### **Listing of Claims:**

1. (Currently amended) A method for finding an optimal set of data association rules in automated data diagnosis of the data characterizing an entity, comprising the steps of:

establishing a computer system for automated data diagnoses;

creating in said computer system a relation R containing the data  $A = (A_1, \dots, A_n, A_{n+1}, \dots, A_{n+m})$ , where  $n, m \geq 1$ , said data characterizing the entity to be diagnosed, said data being represented by outcome attributes  $A_{n+1}, \dots, A_{n+m}$  and by diagnosable attributes  $A_1, \dots, A_n$ , said outcome attributes determining whether the entity is desirable or not, and said diagnosable attributes determining the reason of why the entity is desirable or not;

specifying a selection outcome condition (D) determined by a user which includes strictly outcome attributes selected from said  $A_{n+1}, \dots, A_{n+m}$  attributes;

specifying at least one diagnosable selection condition (C) which includes diagnosable attributes selected by the user from said  $A_1, \dots, A_n$  attributes;

specifying selection condition constraints (S) for said diagnosable selection conditions to meet, said selection condition constraints (S) including minimal acceptable confidence  $\text{conf}(c)$ , minimal acceptance support  $\text{sup}(c)$  and maximum order of said diagnosable selection condition;

specifying a number of fringes of interest,  $F^0, F^1 \dots F^i, F^{i+1}$ ,

wherein

$$F^0 = \{C \mid \text{up}(C) = 0\}$$

$$F^{i+1} = \left\{ C \mid \text{up}(C) \subseteq \bigcup_{j \leq i \wedge C' \in F^j} \right\}$$

and wherein the fringe  $F^0$  represents an optimal set of selection conditions with regard to a combination of respective ones of said specified support, said specified confidence and said specified ~~“simpler-than”~~ simpler-than ordering, the fringe  $F^1$  represents the set of diagnosable selection conditions that are less desirable than the fringe  $F^0$ , and the fringe  $F^{i+1}$  represents the set of diagnosable selection conditions that are less desirable than fringe  $F^i$ ;

computing said optimal fringes  $F^0, F^1, \dots, F^i, F^{i+1}$  ~~[[,]]~~ and ;

computing a compact set of said optimal fringes to eliminate redundant conditions, said compact set representing the optimal set of data association rules; and

storing said compact set of said optimal fringes in a memory of said computer system for retrieval.

2. (Currently amended) The method of Claim 1, further comprising the steps of:

specifying a "~~simpler-than~~" simpler-than ordering ( $\geq$  simpler), including a set of said diagnosable selection conditions (C) which are simpler than a predetermined diagnosable selection condition;

specifying a Data Diagnosis Objective (DDO) by defining components thereof including:

(a) evaluation domain (ED) providing for measurement of quality of said diagnosable selection conditions,

(b) a partial ordering ( $\subseteq$ ) of said evaluation domain, specifying which diagnosable selection condition in said evaluation domain are better than others, and

(c) a mapping function (f) that maps said diagnosable selection conditions to said evaluation domain, such that  $C_1 > C_2 \Rightarrow f(C_2) \subseteq f(C_1)$ ,

wherein  $C_1$  and  $C_2$  are diagnosable selection conditions;

and

specifying a semi-equivalent relation ( $\Delta$ ), on said diagnosable selection conditions to determine similarity thereof.

3. (Original) The method of Claim 1, wherein said entity is a product of a manufacturing process.

4. (Original) The method of Claim 1, wherein said entity is a financial prediction process.

5. (Original) The method of Claim 2, further comprising the step of combining said diagnosable conditions C to form a diagnosable selection condition SC.

6. (Original) The method of Claim 5, further comprising the step of restricting said diagnosable selection conditions to tight diagnosable selection conditions T,

wherein said diagnosable selection condition SC is tight if for each diagnosable condition  $l \leq A_i \leq u$  in said diagnosable selection condition SC,  $(\sigma_{(A_i=l) \wedge SC \wedge D}(R) \neq 0) \wedge (\sigma_{(A_i=u) \wedge SC \wedge D}(R) \neq 0)$  where  $\sigma$  is a relational selection operator, and

wherein  $A_i$  is a diagnosable attribute, and  $l, u, \in \text{dom}(A_i)$  are values defined by the user.

7. (Original) The method of Claim 6, wherein said step of computing said optimal fringes further comprises the steps of:

creating a condition graph by enumerating a set of tight selection conditions  $T$  satisfying said selection condition constraints  $S$ , and  
evaluating support and confidence defined by the user.

8. (Original) The method of Claim 6, wherein if  $l=u$ , then  $A_i = l$ .

9. (Original) The method of Claim 6, wherein  $l=u$  if said  $A_i$  is a non-numeric diagnosable attribute.

10. (Original) The method of Claim 1, wherein said outcome attributes and said diagnosable attributes in said relation  $R$  include numeric or non-numeric attributes.

11. (Original) The method of Claim 6, wherein said  $l \leq A_i \leq v$  is a diagnosable selection condition.

12. (Original) The method of Claim 6, wherein if  $C_1$  and  $C_2$  are diagnosable conditions, then  $C_1 \wedge C_2$  is a diagnosable selection condition.

13. (Original) The method of Claim 6, wherein the confidence  $\text{conf}(c)$  of said diagnosable condition C is defined as

$$\text{conf}(C) = \frac{\text{card}(\sigma_{C \wedge D}(R))}{\text{card}(\sigma_C(R))}.$$

14. (Original) The method of Claim 6, wherein said support  $\text{sup}(c)$  of said diagnosable condition C is defined as

$$\text{sup}(C) = \text{card}(\sigma_{C \wedge D}(R)).$$

15. (Currently amended) The method of Claim 2, wherein said ~~“simpler-  
 than”~~ simpler-than ordering is defined as:

(a)  $C_1$  simpler  $C_2$  if each diagnosable attribute occurring in  $C_1$  also occurs in  $C_2$ ; or

(b)  $C_1 \geq \text{simpler}_2 C_2$  if  $C_1$  has the same or fewer distinct attributes occurring in said  $C_1$ , than in said  $C_2$ ; or

(c)  $C_1 \geq \text{simpler}_3 C_2$  if said  $C_1$  has fewer distinct attributes than said  $C_2$ , or said  $C_1$  has the same number of distinct attributes as  $C_2$  but fewer numeric diagnosable attributes.

16. (Original) The method of Claim 2, further comprising the steps of:

applying standard metrics to said DDO for comparing diagnosable selection conditions including either from a group thereof, consisting of: chi-squared value, confidence, conviction, entropy gain, loplace, lift, gain, gini, and support.

17. (Original) The method of Claim 2, wherein said fringes are defined independent of said DDO.

18. (Original) The method of Claim 17, further comprising the step of applying a plurality of distinct said DDO to said set of fringes avoiding recomputing said set of fringes.

19. (Original) The method of Claim 2, wherein said semi-equivalence relations  $\Delta$  is a distance-based semi-equivalence relation.

20. (Original) The method of Claim 2, wherein said semi-equivalence relation  $\Delta$  is an attribute distance threshold based semi-equivalence relation, the method further comprising the steps of:

defining  $L_{CA}$  and  $U_{CA}$  to be the lower and upper bounds, respectively, of a respective diagnosable attribute A in said diagnosable condition

C, and defining a diagnosable selective condition  $C_1$ , as semi-equivalent to a diagnosable selective condition  $C_2$  if:

- a. the set of diagnosable attributes appearing in said  $C_1$  is equivalent to the set of diagnosable attributes appearing in said  $C_2$ ,
- b. for each numeric diagnosable attribute, A, appearing in said  $C_1$  and said  $C_2$

$$d(L_{C_{1,A}}, L_{C_{2,A}}) \leq \varepsilon_A$$

$$d(U_{C_{1,A}}, U_{C_{2,A}}) \leq \varepsilon_A,$$

where the  $\varepsilon_A$  values are constants that differ based on the diagnosable attribute A; and

- c. for each non-numeric attribute A, appearing in  $C_1$  and  $C_2$ ,

$$L_{C_{1,A}} = L_{C_{2,A}}.$$

21. (Original) The method of Claim 2, further comprising the steps of:

defining a subset SF of said set F of the optimal fringes as said compact representation of said set F with regard to said  $\Delta$ , if:

- a. for each diagnosable selective condition  $sc \in F, \exists sc' \in CF$  such that  $sc \Delta sc'$ ;

- b. if  $sc \in F$  and  $sc \notin CF$ , then  $\exists sc' \in CF$  such that  $sc \Delta sc' \wedge (\neg(f(sc') \subseteq f(sc)) \vee (f(sc') = f(sc)))$ , and



c. there is no strict subset  $CF'$  of  $CF$  satisfying said conditions (a) and (b).

22. (Currently amended) A system for automated data diagnosis, comprising:

a computer system;

means in said computer system for storing data characterizing an entity to be diagnosed,

means for forming a relation  $R$  containing said data, said relation  $R$  containing the data  $A = (A_1, \dots, A_n, A_{n+1}, \dots, A_{n+m})$ , where  $n, m \geq 1$ , the data being represented by outcome attributes  $A_{n+1}, \dots, A_{n+m}$  and diagnosable attributes  $A_1, \dots, A_n$ , said outcome attributes determining whether the entity is desirable or not, and said diagnosable attributes determining the reason of why the entity is desirable or not;

an interface for bi-directional communication between a user and said computer system, the user inputting into said computer system a plurality of selective conditions through said interface $[[, \text{and}]]$  ;

means in said computer system for computing optimal data association rules for said data to be diagnosed, based on said selective conditions,

said selective conditions being characterized in combination, by a confidence, support and simplicity of said selective conditions; and

MR2833-31

Serial Number: 10/774,516

Reply to Office Action dated 29 November 2007

means in said computer system for storing said computed optimal

data association rules.